# *Smart Data Selection*

# **July 2014**

**Tom Young**
**SET Executing Agent**
**412 TENG/ENI**
**(661) 277-1071**
**Email: tommy.young.1@us.af.mil**

**Test Resource Management Center (TRMC)**
**Test & Evaluation/ Science & Technology (T&E/S&T)**
**Spectrum Efficient Technology (SET)**

# REPORT DOCUMENTATION PAGE

| 1. REPORT DATE *(DD-MM-YYYY)* 25-02-2015 | 2. REPORT TYPE Technical Paper | 3. DATES COVERED *(From - To)* 3/13 -- 12/14 |
|---|---|---|

**4. TITLE AND SUBTITLE**

Smart Data Selection

**5a. CONTRACT NUMBER:** W900KK-13-C-0028

**5b. GRANT NUMBER: N/A**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Shannon Wigent & Dr. Andrea Mazzario

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Laulima Systems LLC
1163 Lani Nuu St, Kalaheo, HI 96741-0000

**8. PERFORMING ORGANIZATION REPORT NUMBER**
412TW-PA-14353

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Test Resource Management Center
Test and Evaluation/ Science and Technology
4800 Mark Center Drive, Suite 07J22, Alexandria, VA 22350

**10. SPONSOR/MONITOR'S ACRONYM(S)**
N/A

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** SET 2015-0027

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release A: distribution is unlimited.

**13. SUPPLEMENTARY NOTES**
CA: Air Force Flight Test Center Edwards AFB CA          CC: 012100

**14. ABSTRACT**
The fundamental precept presented in the iNET Concept of Operations, v. 2007.1 is that new telemetry technologies must be created to enable a more flexible approach to testing that includes on-demand access to information acquired on the test article and the ability to reconfigure the telemetry stream definition. …one approach that has not yet been addressed is a concept introduced early within the iNET CONOPS document, that "The dominant inherent nature of TM in DoD testing is sampled time-history data from an ultimately analog world, (which) is not going to change drastically regardless of how data is transmitted to ground. A factor that could change that fact most is the degree to which answers instead of data are obtained on board the test vehicle."
Ultimately, the most effective way of dealing with the exponentially growing gap between the quantities of data generated onboard the test article and the rate at which it is transmitted to ground is to generate answers on board the test article. The Test Resource Management Center (TRMC) Test and Evaluation (T&E) Science and Technology (S&T) Spectrum Efficient Test (SET) Program is sponsoring development of the Smart Data Selection (SDS) system which provides a capability to continually monitor measured data and then select which parameters, or which combination of parameters, to send to ground in a given time interval, based on what is actually happening with the system under test. This paper provides a SDS system update, the results of initial testing, and introduces the PCM compression enhancement that is currently under development. The benefits of this work, in terms of efficient use of spectrum to support T&E are substantial, and could be leveraged by any DoD ranges that execute aeronautical and precision-guided munitions testing.

**15. SUBJECT TERMS**
Spectrum, Aeronautical telemetry, algorithm, bandwidth, Integrated Networked Enhanced Telemetry (iNET), Smart Data Selection (SDS), Pulse Code modulation (PCM)

| 16. SECURITY CLASSIFICATION OF: **Unclassified** | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON 412 TENG/EN (Tech Pubs) |
|---|---|---|---|---|---|
| a. REPORT Unclassified | b. ABSTRACT Unclassified | c. THIS PAGE Unclassified | None | 14 | 19b. TELEPHONE NUMBER *(include area code)* 661-277-8615 |

# SMART DATA SELECTION

**Ms. Shannon Wigent**
**Laulima Systems**

**Dr. Andrea Mazzario**
**Kauai Software Solutions**

## ABSTRACT

The fundamental precept presented in the *iNET Concept of Operations, v. 2007.1* is that new telemetry technologies must be created to enable a more flexible approach to testing that includes on-demand access to information acquired on the test article and the ability to reconfigure the telemetry stream definition. Significant advances have been made in this area but one approach that has not yet been addressed is a concept introduced early within the iNET CONOPS document, that

> *"The dominant inherent nature of TM in DoD testing is sampled time-history data from an ultimately analog world, (which) is not going to change drastically regardless of how data is transmitted to ground. A factor that could change that fact most is the degree to which **answers** instead of data are obtained on board the test vehicle."*

Ultimately, the most effective way of dealing with the exponentially growing gap between the quantities of data generated onboard the test article and the rate at which it is transmitted to ground is to generate answers on board the test article. The Test Resource Management Center (TRMC) Test and Evaluation (T&E) Science and Technology (S&T) Spectrum Efficient Test (SET) Program is sponsoring development of the Smart Data Selection (SDS) system which provides a capability to continually monitor measured data and then select which parameters, or which combination of parameters, to send to ground in a given time interval, based on what is actually happening with the system under test.

The SDS system was initially introduced at ITC in 2013. At that time, the system was under initial development and system testing was not yet completed. This paper provides a SDS system update, the results of initial testing, and introduces the PCM compression enhancement that is currently under development.

The benefits of this work, in terms of efficient use of spectrum to support T&E are substantial, and could be leveraged by any DoD ranges that execute aeronautical and precision-guided munitions testing.

# INTRODUCTION

The fundamental precept presented in the *iNET Concept of Operations, v. 2007.1* is that new telemetry technologies must be created to enable a more flexible approach to testing that includes on-demand access to information acquired on the test article (TA) and the ability to reconfigure the telemetry stream definition. Significant advances have been made in this area by both the iNET and T&E S&T SET programs. One area that has not yet been addressed by these programs is a concept introduced early within the iNET CONOPS document, that

> *"The dominant inherent nature of TM in DoD testing is sampled time-history data from an ultimately analog world, (which) is not going to change drastically regardless of how data is transmitted to ground. A factor that could change that fact most is the degree to which* **answers** *instead of data are obtained on board the test vehicle."*

Ultimately, the most effective way of dealing with the exponentially growing gap between the quantities of data generated onboard the test article and the rate at which it is transmitted to ground – even with advancements brought forth by iNET – is to generate answers on board the test article. The Smart Data Selection (SDS) project is focused on developing a capability that will begin to do just that. The SDS capability continually monitors measured data and then selects which parameters, or which combination of parameters, to send to ground in a given time interval, based on what is actually happening with the system under test. This capability is not intended to replace the existing telemetry paradigm, as some parameters such as safety related data must always be sent to ground. Rather, SDS will augment current approaches to telemetry, and opens the door for an order of magnitude more spectrum efficiency in terms of sending actionable information/sec/Hz to ground.

Fundamentally, our approach uses onboard processing prior to transmission to reduce the volume of data that must be transmitted from the test article. Rather than sending all measured data points, the SDS system applies bandwidth efficient algorithms to selected data, resulting in a significant savings in spectrum. SDS also provides greater operator awareness of system anomalies while ensuring critical data, such as range safety data, is sent to ground continually without interruption.

This paper describes the SDS system design and the results of the initial test phases. SDS is currently adding the PCM compression capability to provide the ability to apply lossless compression algorithms to PCM telemetry data for additional spectrum efficiency. SDS system development is on-going and will be further tests and demonstrations will be conducted during the summer of 2014. The results of these tests and demonstrations will be presented at the International Telemetering Conference in October 2014.

# SYSTEM OVERVIEW

The SDS system design is illustrated in **Error! Reference source not found.** and the major components are described below.
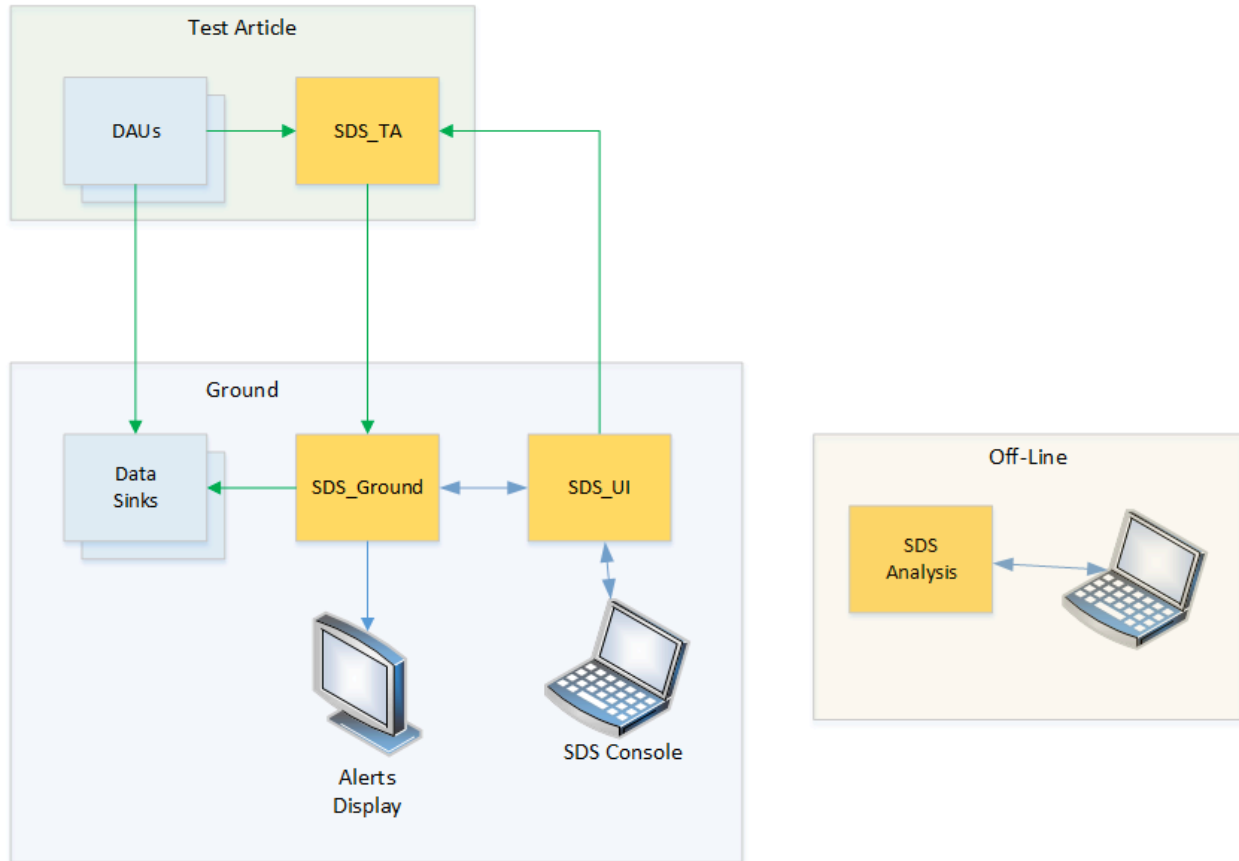


**Figure 1.  SDS System Overview**

The SDS system major components include:
- SDS_TA
  The Test Article (TA) component subscribes to TmNS messages carrying the measurements that SDS will analyze and monitor.  Measurement data that are considered normal are sent to the ground using a bandwidth efficient algorithm.  The user may disable use of the bandwidth efficient algorithm for selected measurements.  Upon detection of an abnormal measurement condition the TA component will generate a TmNS alert message.  The TA component will also begin transmitting the abnormal measurement data without using the bandwidth efficient algorithm.  If a measurement that was behaving abnormally returns to its normal behavior, SDS will generate a corresponding end-of-alert TmNS Message.
- SDS_Ground
  The Ground component is responsible for reconstituting the original TmNS data messages whose measurements have been transmitted using the bandwidth efficient algorithm and for publishing them to the ground network.  It is also responsible for

displaying the alerts generated by the TA component on its Alert Display, which also allows the user to mark a displayed alert as a "false alert".

- SDS_UI
  The UI component is responsible for accepting user commands that control the compression of packages and the handling of individual alert messages.
- SDS_Analysis
  The Analyze component analyzes recorded data files to determine behavioral patterns as well as utilizes user-define behavioral criteria. The analysis of recorded data and the user-defined criteria are used to establish what is considered normal and abnormal behavior or what data is of interest for real-time observation and analysis.

## ERROR BOUND EXTRAPOLATION ALGORITHMS

The error bound extrapolation (EBE) algorithms implemented in SDS are based on existing extrapolation algorithms. Rather than transmitting individual measurement values, the SDS_TA transmits extrapolation parameters. The SDS_Ground uses these parameters to calculate and publish the measurement values with the required frequency. At the same time, the SDS_TA monitors the error between the extrapolated and the actual measurements. Whenever the error value exceeds a user defined threshold, the TA SDS calculates and sends new extrapolation parameters to the SDS_Ground. More specifically, the SDS_TA uses a first algorithm to smooth the data as it receives them, and then calculates the extrapolation parameters for the smoothed data. These extrapolation parameters are sent to the SDS_Ground which uses them to publish the reconstructed measurements. At the same time, the SDS_TA uses the same extrapolation parameters to compare the "current extrapolated value" (which is the value used at that time by the SDS_Ground) with the "current smoothed value". If the difference exceeds the error threshold specified for that measurement, a new set of extrapolation parameters (i.e., an "extrapolation reset") is sent to the SDS_Ground.

### Exponential Smoothing

The SDS_TA uses exponential smoothing to calculate the smoothed values of the measurements. There are multiple levels of smoothing, but tests performed on real measurement data revealed that the best results are obtained by using single or double exponential smoothing.

*Single Exponential Smoothing*

In the single exponential smoothing algorithm, the smoothed value sv[i] of a measurement m is calculated using the formula:

$$sv[i] = \alpha * m[i] + (1 - \alpha) * sv[i-1]$$

The initial value sv[0] is usually set to 0, and $\alpha$ is a chosen value between 0 and 1. The value of $\alpha$ controls the degree of smoothing: an $\alpha$ close to 1 does little smoothing (the smoothed value follows the current value relatively closely) while a small $\alpha$ does more smoothing (the smoothed value is very stable).

*Double Exponential Smoothing*

In the double exponential smoothing algorithm, the smoothed value sv[i] of a measurement m is calculated using the formula:

$$sv[i] = \alpha*m[i] + (1 - \alpha)*(sv[i-1] + t[i-1])$$
$$t[i] = \beta * (sv[i] - sv[i-1]) + (1 - \beta)*t[i-1]$$

The initial values sv[0] and t[0] are usually set to 0, and $\alpha$ and $\beta$ are chosen value between 0 and 1. Their values control the degree of smoothing: values close to 1 do little smoothing (the smoothed value follows the current value relatively closely) while small values do more smoothing (the smoothed value is very stable).

**Choice of the Smoothing Algorithm and of the Smoothing Parameters**

The SDS_Analysis tool will evaluate multiple recordings of a measurement and recommend the best smoothing algorithm (simple or double exponential smoothing) and the best parameter ($\alpha$ and possibly $\beta$) to use.

**Extrapolation**

Extrapolation parameters result from the exponential smoothing algorithms. For single exponential smoothing, the extrapolated value ev[i+n] is derived from the smoothed value sv[i] using the formula:

$$ev[i+n] = sv[i]$$

In other words, the extrapolated value is a constant. This constant is adjusted when the difference to the actual smoothed value exceeds the user specified threshold.

For double exponential smoothing, the extrapolated value ev[i+n] is derived from the smoothed value sv[i] and the trend t[i] using the formula:

$$ev[i+n] = sv[i] + n*t[i]$$

In other words, we have a linear extrapolation. The extrapolation parameters (sv[i[ and t[i]) are adjusted when the difference to the actual smoothed value exceeds the user specified threshold.

**Reset of the Smoothing Function**

The SDS_TA also monitors the difference between the current measurement and the current smoothed measurement value. If this error exceeds a given threshold, usually related to the intrinsic error and variability of the measurement, it is deemed to be a sudden change of the measurement itself that cannot be followed by the smoothing algorithm. In this event, the smoothing function is reset to reflect the actual value of the measurement. Any reset of the smoothing function results in a reset of the extrapolation parameters, which are then transmitted to the SDS_Ground.

**Error Threshold**

The acceptable error threshold for the extrapolation depends on one hand on the user's requirements, and on the other hand on the intrinsic accuracy of the measurement. In other words, the acceptable error should be larger than the intrinsic fluctuation of the measurements.

The SDS can accept an absolute error threshold, or a multiple of the intrinsic measurement error. In this last case, if the intrinsic measurement error is not provided, the SDS will initially estimate it as the error between the measurements and their smoothed value.

**Compression Examples**

*Thermocouple A*

This example uses a recording of a thermocouple measurement. The represented values are the raw telemetry data, before application of the Engineering Unit conversions. Frequency of the measurement is 98.04 Hz and there are ~45000 measurements with values between 32500 and 41000.

In the following graphs, red dots represent smoothing resets and green dots represents extrapolation resets.  In the enlargements, the blue dots represent the raw values, the black line represents the smoothed value, and the green line represents the extrapolated value.

Example 1: Maximum error < 5 times intrinsic error

In this example, as shown in Figure 2, we generated an extrapolation with a very small error threshold. 5 times the intrinsic error results in a maximum error between the measurement and the extrapolated value of $\sim 5$, which results in an error <= 0.01%.

This measurement uses double exponential smoothing with $\alpha = 0.1$ and $\beta = 0.1$.
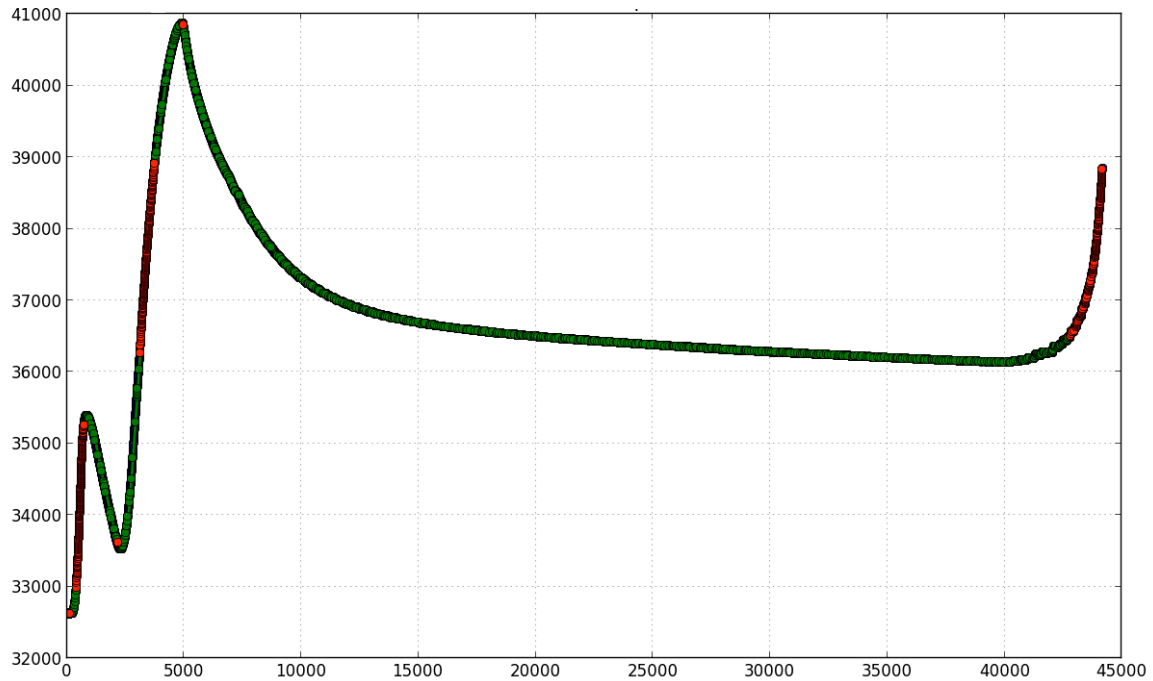
**Figure 2.  Example 1**

To transmit this measurement to the ground the SDS needs 1001 resets to represent the 44091 measurements. Given some overhead and the need to transmit 2 values per reset, the transmission cost of a reset is ~ 3 measurements. At 2 bytes per measurement, we can estimate the amount of data transmitted to the ground:

SDS compressed data:        2 * 3 * 1,001 = 6,006
Raw Measurements:        2 * 44,091 = 88,192

In other words, the SDS needs less that 7% of the bandwidth required to transmit the original data. One could think that this is a very special case, with the parameters finely tuned to obtain the best result, but looking close at the graph we can show that this is not the case.

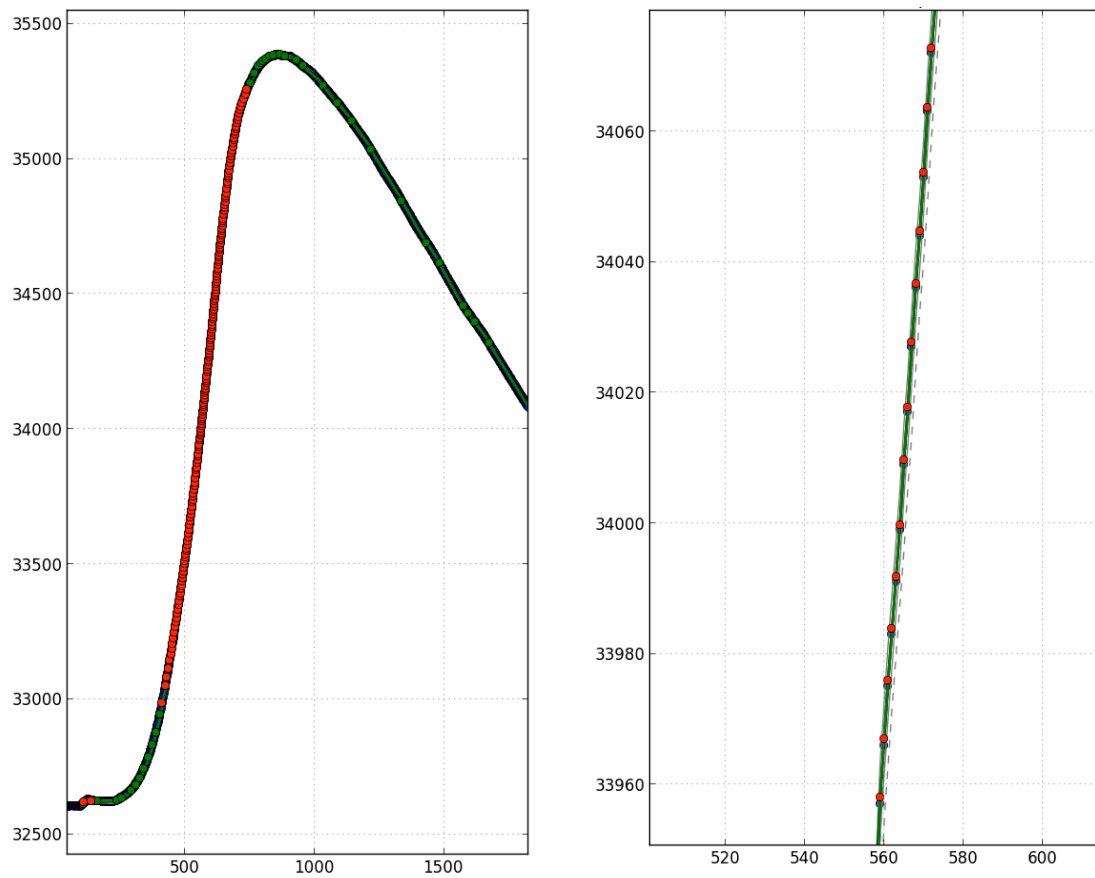In Figure 3, we enlarge the initial, very steep rise:

6

**Figure 3.  Example 1 Enlarged**

What happened here is that the rise of the function is so steep that the smoothing cannot follow, and practically every measurement causes a smoothing reset, which is not an optimal situation. Near the top of the curve, the smoothing function can finally keep up which results in less frequent extrapolation resets, especially in the descending arm, as shown in Figure 4.
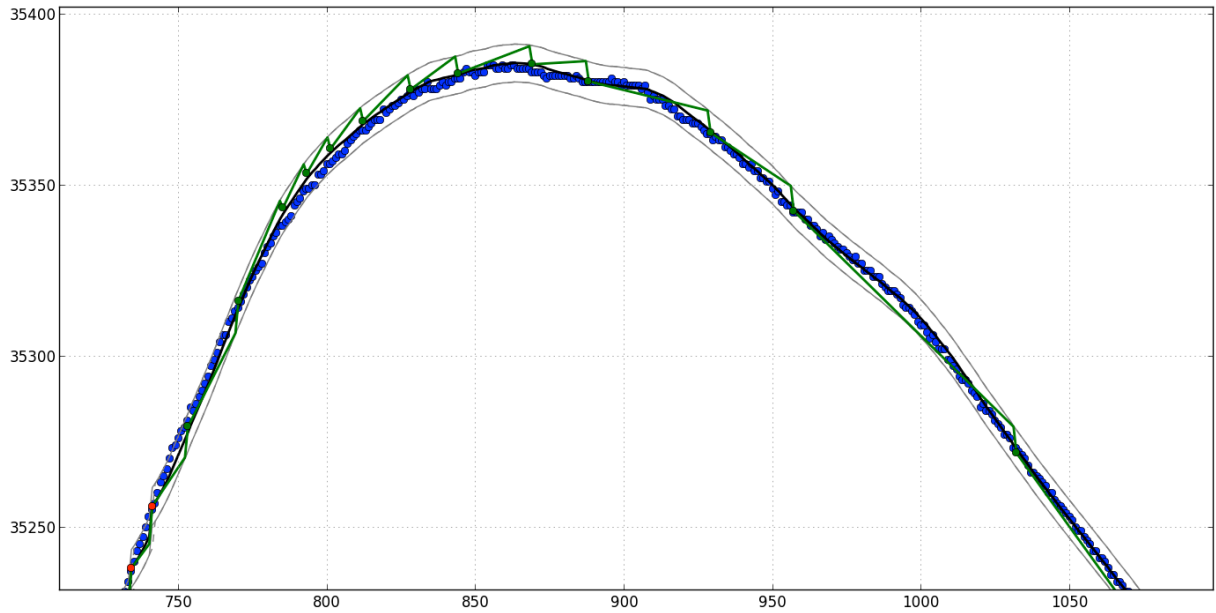
**Figure 4.  Example 1 Enlarged**

It is also interesting to notice that the almost flat area in the center of the original graph, between 15,000 and 40,000, when enlarged is not flat, but the extrapolation algorithm still manages to follow it quite accurately, Figure 5.
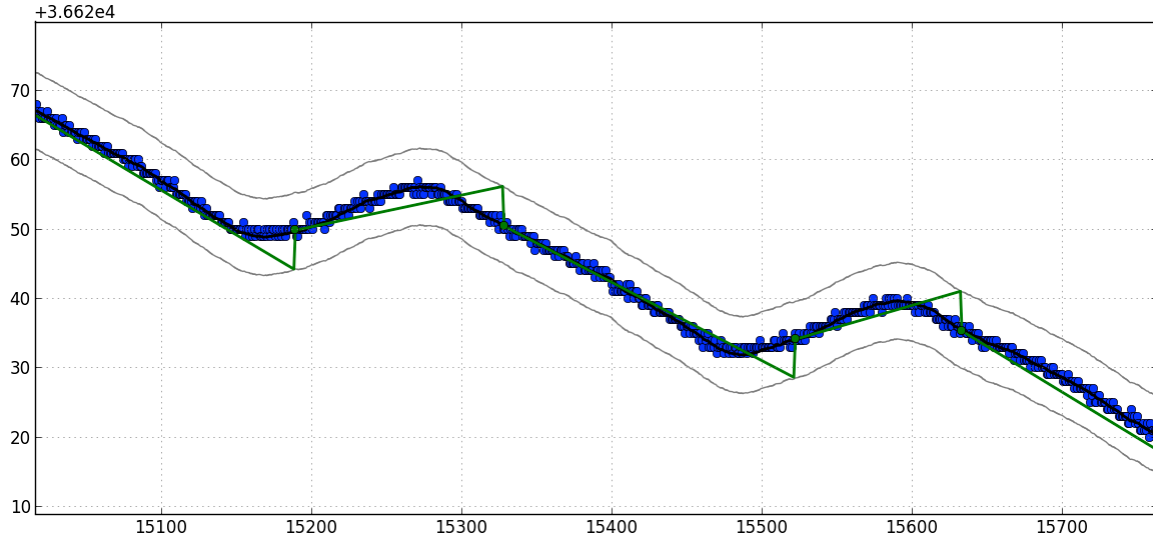


**Figure 5.  Example 1 Enlarged**

Example 2: Maximum error < 10 times intrinsic error

If a somewhat smaller accuracy of the extrapolated data is necessary, it is possible to achieve even greater bandwidth savings.  10 times the intrinsic error results in a maximum error between the measurement and the extrapolated value of ~ 10, which results in an error <= 0.02%.

8

Keeping all the other parameters the same as in the previous example, we now need only 446 resets to represent the 44091 measurements. We perform the same calculations as in the previous case:

SDS data: 2 * 3 * 446 = 2,676
Raw Measurements: 2 * 44,091 = 88,192

We see that in this case the SDS needs only 3% of the bandwidth required to send the original data. The behavior of the extrapolation is very similar to the previous example, only the allowed error is larger, so that resets can be further apart, Figure 6.
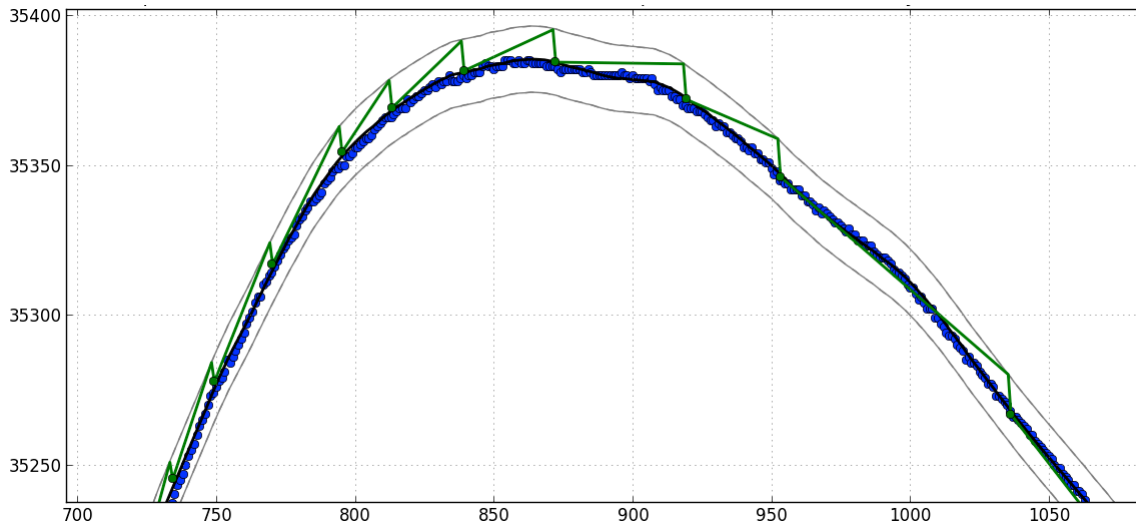


**Figure 6.  Example 2 Enlarged**

### *Thermocouple B*

We ran a similar test with a different thermocouple from the same test article.  This thermocouple also had 44091 measurements with a rate of 98.04 Hz.

Example 3: Maximum error < 5 times intrinsic error

In this example, Figure 7, we generated an extrapolation with a very small error threshold. 5 times the intrinsic error results in a maximum error between the measurement and the extrapolated value of ~ 5, which results in an error <= 0.01%.

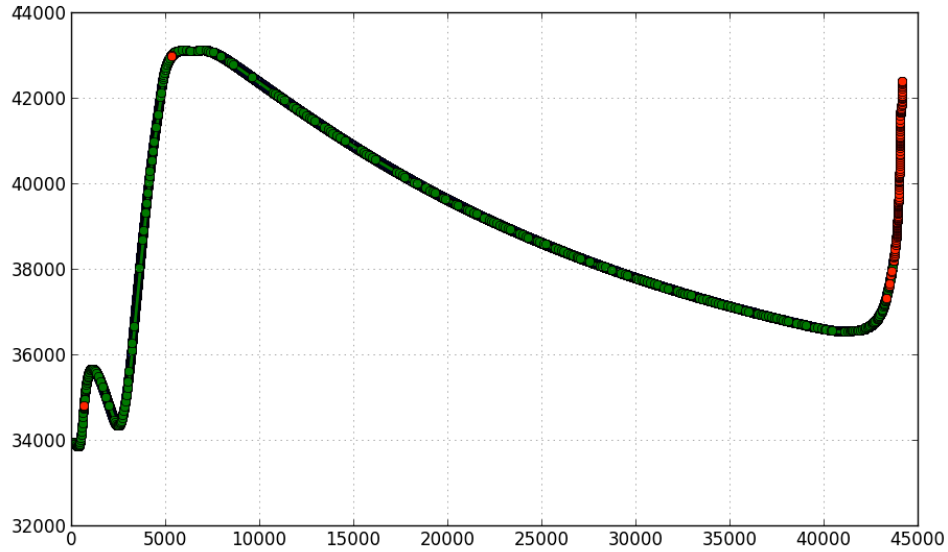This measurement uses double exponential smoothing with $\alpha = 0.1$ and $\beta = 0.1$.

9

**Figure 7. Example 3**

To transmit this measurement to the ground the SDS needs 492 resets to represent the 44091 measurements. Given some overhead and the need to transmit 2 values per reset, the transmission cost of a reset is ~ 3 measurements. At 2 bytes per measurement, we can estimate the amount of data transmitted to the ground:

SDS data: $2 * 3 * 492 = 2,952$
Raw Measurements: $2 * 44,091 = 88,192$

In other words, the SDS needs only 3.4% of the bandwidth required to send the original data. The extrapolation algorithm applied to Example 3 results in greater savings than Example 1 due to the nature of the curve. In Example 3, the extrapolation algorithm requires less resets because the measurements between 10,000-40,000, Figure 8, is more smooth than in Example 1.
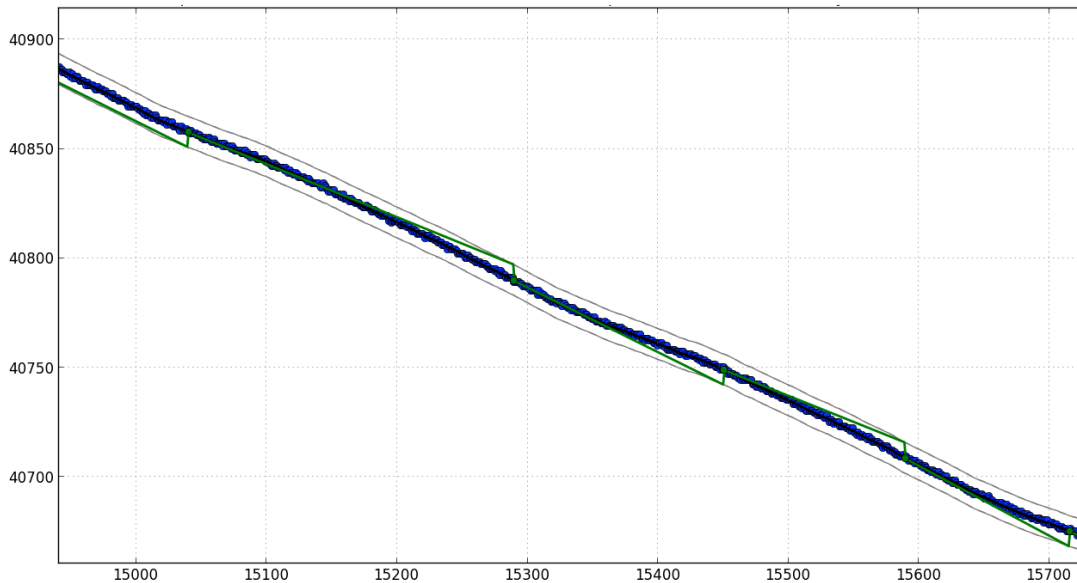


**Figure 8. Example 3 Enlarged**

10

Example 4: Maximum error < 10 times intrinsic error

If a smaller accuracy of the extrapolated data is necessary, it is possible to achieve even greater bandwidth saving is possible. 10 times the intrinsic error results in a maximum error between the measurement and the extrapolated value of ~ 10, which results in an error <= 0.02%.

Keeping all the other parameters the same as in the previous example, we now need only 294 resets to represent the 44091 measurements. We perform the same calculations as in the previous case:

| | |
|---|---|
| SDS data: | 2 * 3 * 294 = 1,764 |
| Raw Measurements: | 2 * 44,091 = 88,192 |

We see that in this case the SDS needs only 2% of the bandwidth required to send the original data.

## CONCLUSIONS

The addition of the SDS capability onboard the test article provides bandwidth savings and increased spectrum efficiency to the T&E community. Rather than sending all measured data points, the SDS system applies error bound extrapolation algorithms to selected data, resulting in a significant savings in spectrum. SDS also provides greater operator awareness of system anomalies while ensuring critical data, such as range safety data, is sent to ground continually without interruption.

The benefits of this work, in terms of efficient use of spectrum to support T&E are substantial, and could be leveraged by any DoD ranges that execute aeronautical and precision-guided munitions testing.

## ACKNOWLEDGEMENTS